

Data-driven Two-step Day-ahead Electricity Price Forecasting Considering Price Spikes

Shengyuan Liu, Yicheng Jiang, Zhenzhi Lin, *Member, IEEE*, Fushuan Wen, *Fellow, IEEE*, Yi Ding, *Member, IEEE*, and Li Yang, *Member, IEEE*

Abstract—In the electricity market environment, electricity price forecasting plays an essential role in the decision-making process of a power generation company, especially in developing the optimal bidding strategy for maximizing revenues. Hence, it is necessary for a power generation company to develop an accurate electricity price forecasting algorithm. Given this background, this paper proposes a two-step day-ahead electricity price forecasting algorithm based on the weighted K -nearest neighborhood (WKNN) method and the Gaussian process regression (GPR) approach. In the first step, several predictors, i.e., operation indicators, are presented and the WKNN method is employed to detect the day-ahead price spike based on these indicators. In the second step, the outputs of the first step are regarded as a new predictor, and it is utilized together with the operation indicators to accurately forecast the electricity price based on the GPR approach. The proposed algorithm is verified by actual market data in Pennsylvania-New Jersey-Maryland Interconnection (PJM), and comparisons between this algorithm and existing ones are also made to demonstrate the effectiveness of the proposed algorithm. Simulation results show that the proposed algorithm can attain accurate price forecasting results even with several price spikes in historical electricity price data.

Index Terms—Electricity market, electricity price forecasting, price spike, weighted K -nearest neighborhood (WKNN), Gaussian process regression (GPR).

I. INTRODUCTION

IN a competitive electricity market, electricity price is one of the most essential factors that should be taken into full consideration by a power generation company since it plays a very important role in building its optimal bidding strategy to maximize the economic benefits. Hence, it is of great sig-

nificance for a power generation company to accurately forecast the electricity price for participating in the spot market. The fluctuation of electricity price depends on the combined effect of many subjective as well as objective factors, involving complicated randomness in its evolution process [1]–[3]. Furthermore, the random occurrence of price spikes presents a huge challenge since the variations of electricity price in this case does not follow the normal pattern, which leads to more difficulties for accurate electricity price forecasting. For example, more than 2.7 million residents lost the power supply due to the unusual cold spell in February 2021, and the electricity price was boosted up to 9000 \$/MWh. Therefore, it is essential for a power generation company to propose an electricity price forecasting algorithm that could properly address the impacts of price spikes for achieving good forecasting performance.

So far, several approaches or models have been presented for forecasting the electricity price in day-ahead electricity markets. In [4], nonparametric possibility distribution models for the electricity price are established through convex optimization and used for electricity price forecasting. In [5], optimal power flow (OPF) is applied in the supply-demand matching process, and the structure of an electricity market is formalized and the nodal prices are forecasted by a holistic approach. In [6], several deep learning approaches for forecasting spot electricity prices are discussed, and it is concluded that the deep neural network (DNN), long short-term memory (LSTM) network, and gated recurrent unit (GRU) network can achieve better performance than the convolution neural network (CNN). In [7], additive regression models are combined with functional covariates to forecast the electricity price. In [8], bagging and random forests based on the regression tree are used for price forecasting. Both approaches in [7], [8] are tested in the Spanish electricity market to demonstrate their effectiveness.

In [9], grey correlation analysis (GCA) and kernel principal component analysis (KPCA) are respectively utilized to select the features corresponding to electricity price and achieve dimensional reduction, and the support vector machine (SVM) based on differential evolution (DE) is employed to forecast electricity price. In [10], the similar day (SD) approach and artificial neural network (ANN) are combined to forecast day-ahead electricity price in the Pennsylvania-New Jersey-Maryland Interconnection (PJM) electricity market. Similarly, ANN is also used in [11] with wavelet

Manuscript received: March 22, 2021; revised: October 7, 2021; accepted: December 28, 2021. Date of CrossCheck: December 28, 2021. Date of online publication: February 4, 2022.

This work was supported by National Natural Science Foundation of China (No. 52077195) and Zhejiang University Academic Award for Outstanding Doctoral Candidates (No. 202022).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

S. Liu, Z. Lin (corresponding author), F. Wen, Y. Ding, and L. Yang are with the School of Electrical Engineering, Zhejiang University, Hangzhou 310027, China, and Z. Lin is also with the School of Electrical Engineering, Shandong University, Jinan 250061, China (e-mail: eelsy@zju.edu.cn; linzhenzhi@zju.edu.cn; fushuan.wen@gmail.com; yiding@zju.edu.cn; eeyangli@zju.edu.cn).

Y. Jiang is with the State Grid Zhejiang Electric Power Corporation, Hangzhou 310007, China, on leave from Zhejiang University, Hangzhou 310027, China (e-mail: marissajyc@gmail.com).

DOI: 10.35833/MPCE.2021.000196



analysis based on the Mexican hat wavelet to achieve higher forecasting accuracy. Considering the limitations of ANN, e. g., slow training speed and local optimality, extreme learning machine (ELM) and bootstrapping are employed in [12] and show potential capabilities for online market price forecasting.

In [13] and [14], the concept of price perturbation intervals is introduced, with the price forecasted in a range subject to a given reliability level rather than a single value. Besides, ELM and the non-dominated sorting genetic algorithm II (NSGA-II) are utilized together to find the Pareto optimal perturbation intervals for electricity price. In [15], the relevance vector machine (RVM) is presented for price forecasting with regression coefficient evaluation based on the micro-genetic algorithm. It should be noted that it is important to extract and select appropriate features as the input data for electricity price forecasting. In [16], the concept of interaction of price feature selection is addressed and measured by information-theoretic criteria (ITC). Then, the hybrid filter-wrapper approach with relevancy, redundancy, and interaction among candidate input data considered is used to find the most appropriate features for forecasting. In [17], a weight nearest neighborhood (WNN) based approach is proposed, and the electricity price is forecasted by the linearly weighted combinations of the actual former prices. In [18], the Elman network is utilized for electricity price forecasting considering its irregular fluctuation, which achieves better performance than autoregressive integrated moving average (ARIMA) and WNN based approaches. In [19], a stacked nosing autoencoder (SDA) model is extended to forecast electricity price, whose effectiveness is further validated in industrial applications.

In [20], normal and random spikes of electricity spot price are handled by the autoregressive time-varying model and kernel regression, respectively. In [21], an intra-hour rolling horizon framework is presented, which can recognize the spike and large variation of electricity prices. In addition, a review of state-of-the-art techniques about electricity price forecasting is given in [22], which aims to illustrate the principle of the solutions and their strengths and weaknesses. Furthermore, a review of probabilistic forecasting technique that aims to predict quantiles or the whole distribution is also represented in [23] for forecasting electricity price reasonably. To study the influence of price spike on electricity price forecasting, the kernel SVM technique is utilized in [24] to determine whether the price is in normal ranges or is a spike so as to forecast the price more accurately with the SVM regression or Bayesian classifier with benefit maximization (BCBM) [25]. In [24] and [25], existence is introduced as a binary indicator for each time to indicate whether a price spike occurs in the past 24 hours and it is confirmed from statistical analysis on the Queensland (QLD) market that spikes tend to occur together over several hours but no longer than a day. In [26], a novel multivariate dynamic model for price forecasting is proposed and the copula representation is utilized for characterizing the joint distribution of electricity price so as to achieve a better result. In [27], the forecasting of price spikes in the Ontario market is dis-

cussed, and different neural networks are utilized to forecast the electricity price with and without price spikes.

However, there are still several limitations for these proposed approaches or models and they can be summarized into six categories. ① The nonlinearity of market price is not considered in the employed time-series-based techniques and the criteria to determine the window length is not addressed, which would lead to less accurate results, e.g., [4], [5], and [7]. ② Electricity price spikes are not considered, which could lead to large forecasting errors in real market operation, e.g., [4]-[19]. ③ The nonlinearity of market price is considered in the employed multi-layer feed-forward neural network with the backpropagation training algorithm, but the performance is inferior to those of deep learning approaches, e.g., [8], [10], and [11]. ④ Accurate market price forecasting is achieved by several deep learning approaches, but it is hard to determine the hyper-parameters which have significant impacts on the final performance and hence on potential practical applications, e.g., [6]. ⑤ Price spikes are considered separately from normal price patterns in the forecasting process but how to detect the existence of price spikes is not addressed in detail, e.g., [20]. ⑥ Price spikes can be detected, but the forecasting accuracy can be further improved by employing available advanced methods such as sophisticated machine learning approaches, e.g., [9], [21]-[27].

In light of the above considerations, this paper aims to propose a two-step day-ahead electricity price forecasting algorithm. In the first step, the occurrence of price spikes is determined in advance by the weighted K -nearest neighborhood (WKNN) method considering several spike-related indicators, which can be regarded as a classification problem. In the second step, the situation indicator of the price, i.e., 0 for the normal situation and 1 for the spike situation, is added as a new indicator and the electricity price is forecasted based on the Gaussian process regression (GPR) approach, which can be regarded as a regression problem. The main contributions of this paper can be summarized as follows.

1) Compared with previous studies, several new operation indicators including prior knowledge and recorded historical ones are presented and they are combined to detect the occurrence of price spikes in advance with the help of the WKNN method. Besides, the importance degree, i. e., weight, of data is also considered according to the variation characteristics of electricity price, which can detect the price spikes more accurately.

2) With the characteristics of electricity price data considered, the GPR approach with the exponential kernel function and hyper-parameters determined by Bayes estimation is initially employed for electricity price forecasting. Case studies show that it can achieve less forecasting error compared with other price forecasting methods.

3) With the help of the proposed two-step day-ahead electricity price forecasting algorithm, it is no longer needed to forecast the price in normal and spike situations separately. Whether price spikes will occur is denoted as one of the input parameters and is utilized together with the presented operation indicators in the second step. Therefore, the proposed algorithm can distinguish these two situations automatically

and only one GPR-based forecasting model rather than two is required, which simplifies the forecasting process.

II. ELECTRICITY PRICE SPIKE DETECTION BASED ON WKNN METHOD

As mentioned above, price forecasting largely depends on regression analysis, i.e., a typical kind of predictive modeling technology. To be specific, based on the analysis of historical data, a certain relationship among variables can be detected and thus a time series model can be built to forecast the future market prices. In a price series, price spikes refer to those that are exceptionally higher or lower than normal ones, which are usually caused by a short-term imbalance between power supply and demand. For example, on August 2019 and February 2021, electricity prices soared in the wholesale electricity market of Texas, USA exceeded 9000 \$/MWh [28]. In fact, the price spikes are outlier points in the price series, which can greatly influence the performance of price forecasting with regression approaches. To make up for this deficiency and reduce the impact of price spikes on price forecasting as much as possible, the electricity price spike detection method based on the WKNN-based method is proposed in this section.

Before employing the proposed WKNN-based method, several predictors, i.e., operation indicators, associated with the occurrence of price spikes, are proposed based on the information disclosure of an electricity market. The operation indicators can be divided into two types, i.e., prior knowledge indicators associated with the forecasted day shown in Table I and recorded historical indicators shown in Table II. It can be observed that all the indicators in Table I can be obtained before the day of price forecasting, although they may not be exactly the same as the actual data during the next day's operation. Specifically, the 1st and 2nd indicators in Table I are the date information which is known in advance; the 3rd and 7th indicators are forecasted by power generation companies and can be obtained one day in advance; and the 4th, 5th, and 6th indicators are scheduled by power generation companies and can also be obtained one day in advance. All the indicators listed in Table II can be determined based on actual recorded data, but only the past, i.e., historical, values of them can be utilized for electricity price spike and price forecasting in current stage. Considering the inherent periodicity of electricity market operation [29], the week-ahead and day-ahead indicators are selected. It is worth mentioning that: ① h^{EC} and h^{EM} represent the total amounts of economic and emergency megawatts offered in the energy market based on offers of power generators, respectively; and ② h^{FGO} is forecasted by PJM based on available information. More details about the indicators listed in Table I and Table II can be found in [30]-[35].

These operation indicators concerning both power generation and consumption have certain impacts on the occurrence of the price spike. The philosophy of selecting the operation indicators can be summarized as follows. To consider the supply-demand relationship of electricity, the indicators h^{FL} and h^{FSG} are presented. To consider the cost of generation and spinning reserve, the indicators h^{EC} and h^{EM} are pre-

sented. To consider the weather variation in different months and user behaviors, the indicators h^M and h^T are presented. To reflect the outage risks, the indicator h^{FGO} is presented. In fact, electricity price series contain several nonstationary features such as trends, changes in level and slope, and seasonality. These features are often important and have impact on the parts of the price signal, which should be considered. On the one hand, h^M and h^T can partly characterize the nonstationary features and the changes of trends and seasonality; on the other hand, the recorded historical indicators $h^{AL,d-7}$, $h^{AL,d-1}$, $h^{LG,d-7}$, $h^{LG,d-1}$, $h^{TL,d-7}$, and $h^{TL,d-1}$ can also provide periodic and actual information that contributes to price spike detection.

TABLE I
PRIOR KNOWLEDGE INDICATORS ASSOCIATED WITH FORECASTED DAY

No.	Prior knowledge data	Reference	Indicator
1	Month		h^M
2	Time		h^T
3	Forecasted load	[30]	h^{FL}
4	Scheduled total generation output	[31]	h^{FSG}
5	The maximum economic generation capacity	[32]	h^{EC}
6	The maximum emergency generation capacity	[32]	h^{EM}
7	Forecasted generation outage	[33]	h^{FGO}

TABLE II
RECORDED HISTORICAL INDICATORS

No.	Historical data	Reference	Indicator
8	Week-ahead actual load	[34]	$h^{AL,d-7}$
9	Day-ahead actual load	[34]	$h^{AL,d-1}$
10	Week-ahead total generation output	[35]	$h^{LG,d-7}$
11	Day-ahead total generation output	[35]	$h^{LG,d-1}$
12	Week-ahead total power loss	[35]	$h^{TL,d-7}$
13	Day-ahead total power loss	[35]	$h^{TL,d-1}$

It should be mentioned that the price spikes can be caused by extreme weather, transmission congestion, forced generator or line outage, tight reserve capacity, and a high concentration ratio of reserve capacity. Although indirectly, these factors are considered in the presented indicators. For example, the extreme weather and forced generator or line outage can be reflected by the indicators h^{EM} and h^{FGO} . Similarly, transmission congestion, tight reserve capacity, and high concentration ratio of reserve capacity can be reflected by the indicators h^{FSG} and h^{EC} . Therefore, these physical factors are considered indirectly indeed. The reason why they are not taken into consideration directly is that these physical factors are not quantified as data in the current database. Thus, this paper considers these factors through several operation indicators for detecting price spikes.

Hence, these operation indicators are utilized as the input data of the WKNN classifier. The feature of the i^{th} price data can be denoted by an indicator vector x_i as:

$$x_i = [h_i^{FL}, h_i^{FSG}, h_i^{EC}, h_i^{EM}, h_i^M, h_i^T, h_i^{FGO}, h_i^{AL,d-7}, h_i^{AL,d-1}, h_i^{LG,d-7}, h_i^{LG,d-1}, h_i^{TL,d-7}, h_i^{TL,d-1}] \quad (1)$$

Besides, the target data of the WKNN classifier can be obtained according to the thresholds determined in [24] as:

$$P_{\text{thres}}^{\text{H}} = \mu_p + 2\sigma_p \quad (2)$$

$$P_{\text{thres}}^{\text{L}} = \mu_p - 2\sigma_p \quad (3)$$

where $P_{\text{thres}}^{\text{H}}$ and $P_{\text{thres}}^{\text{L}}$ are the upper and lower thresholds for defining price spikes, respectively; and μ_p and σ_p are the mean value and standard deviation of historical electricity prices, respectively. If the market price P_i is beyond the range of $[P_{\text{thres}}^{\text{L}}, P_{\text{thres}}^{\text{H}}]$, it is then regarded as the price spike ($y_i = 1$); otherwise, it is regarded as the normal price ($y_i = 0$). Thus, the input and target data of the WKNN classifier for price spike detection are all settled. The WKNN classifier of price spikes f_{WKNN} can be represented as:

$$y_{\text{new}} = f_{\text{WKNN}}(\mathbf{x}_{\text{new}}(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_{N_{\text{train}}}, y_{N_{\text{train}}})) \quad (4)$$

where $y_{\text{new}} \in \{0, 1\}$ and \mathbf{x}_{new} are the label and indicator vector of data point to be classified, respectively; and N_{train} is the number of training data samples, which is related to the size of the data set. In fact, the WKNN method classifies the samples, i.e., prices at different time, by measuring the distances among their several features, i.e., the operation indicators associated with price spikes. The motivations of using WKNN method are: ① K -nearest neighborhood (KNN) and WKNN methods are online technologies where new data can be added directly to the data set without retraining; ② KNN and WKNN methods can be used for nonlinear classification; and ③ the computation complexity of KNN and WKNN methods is $O(n)$, which can be solved quickly. Besides, the most essential improvement of WKNN method compared with KNN method is that the voting process of WKNN method is weighted. It is widely accepted that a gradual increase/decrease in electricity price is normal, but an abrupt change should be recognized as a spike. Therefore, the KNN method may misclassify the normal data points of price into the spike ones while the WKNN method will not. This improvement is based on the idea that the data point which is particularly near the unclassified data point should be given much more considerations, i.e., a larger weight. The basic idea of WKNN method is that if the weighted majority of the K most similar samples, i.e., the K nearest neighbors in the feature space, of a sample belong to a certain class, then the sample will also be determined as an element of this class. It should be noted that the parameter K in WKNN method is an integer smaller than 20, and it can be given in advance by experience or determined through the WKNN training process automatically [36]. To illustrate the idea of the WKNN method for price spike detection, a schematic diagram is given in Fig. 1.

The first process is to look at the data, i.e., to find the locations of currently detecting price data with unknown class and past detected price data with individual but known class labels, i.e., $y_i = 0$ or $y_i = 1$, in feature space. For example, the black point represents the price data to be classified, the red points represent the price spike data in the past, and the green points represent the normal price data in the past, respectively.

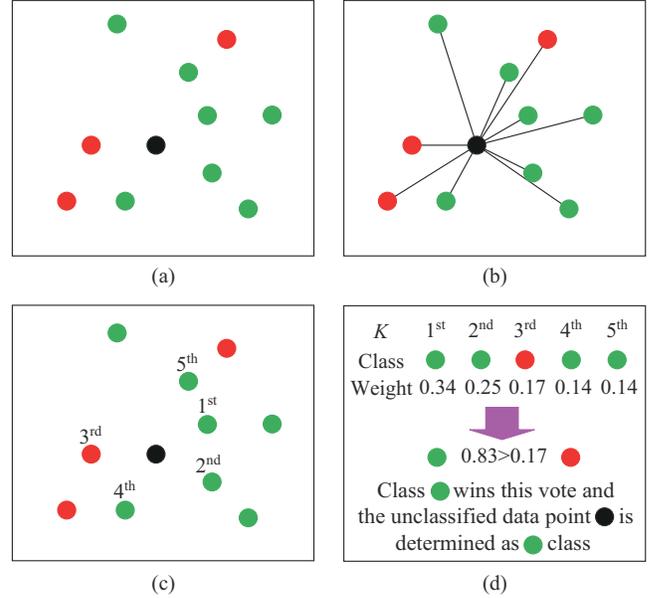


Fig. 1. Schematic diagram of WKNN method. (a) Process 1: look at data. (b) Process 2: calculate distances. (c) Process 3: find neighbors. (d) Process 4: vote by weight majority.

The second process is to calculate the distances between the unclassified price data and all the other classified data one by one, i.e., the distances between the black point and other points.

The third process is to find the neighbors of the black point by ranking the other points in ascending order. For example, the point with the shortest distance away from the black one is the 1st nearest neighbor, the point with the second shortest distance away from the black one is the 2nd nearest neighbor, and the rest can be finished in the same manner. Thus, the K nearest neighbors of \mathbf{x}_{new} are obtained and they can be denoted as a set ϕ_K .

The final process is to vote the class of unclassified data points by the weighted majority. The more recent the price data are, the larger the weights will be set, vice versa. In WKNN method, the weights w_1, w_2, \dots, w_K of different data points are determined by the kernel function. More discussions about the distance and weight determinations can be found in [37]. Thus, the class label of \mathbf{x}_{new} can be determined by:

$$y_{\text{new}} = \arg \max_{\vartheta} \sum_{(\mathbf{x}_i, y_i) \in \phi_K} w_i I(\vartheta = y_i) \quad (5)$$

where $I(\cdot)$ is the indicator function, which equals to 1 if $\vartheta = y_i$ is true and equals to 0 if $\vartheta = y_i$ is false.

The case in Fig. 1 is taken as an example to further illustrate the above-mentioned process. There are 4 points belonging to the normal price data set while one point belonging to spike price data if $K=5$ is given. Assume that the weights of these points determined are $w_1=0.34$, $w_2=0.25$, $w_3=0.17$, $w_4=0.14$, and $w_5=0.10$, respectively, then the unclassified data point will be classified as the normal one, since $(0.34 + 0.25 + 0.14 + 0.10) \times 1 > 0.17 \times 1$, and thus the normal class wins the vote in this case.

III. ELECTRICITY PRICE FORECASTING BASED ON GPR APPROACH

Once the price spike in a given price series has been detected, the indicator y can be used as a new indicator for electricity price forecasting apart from the previously defined indicator x_i . Therefore, the electricity price forecasting can be represented as a regression analysis problem and formulated as:

$$z_{\text{new}} = f_r(\mathbf{v}_{\text{new}} | (\mathbf{v}_1, z_1), (\mathbf{v}_2, z_2), \dots, (\mathbf{v}_{N_{\text{train}}}, z_{N_{\text{train}}})) \quad (6)$$

$$\mathbf{v}_i = (\mathbf{x}_i, y_i) \quad (7)$$

where $z_{\text{new}} \in \mathbf{R}$ and \mathbf{v}_{new} are the electricity price to be forecasted and corresponding input data vector, respectively; z_i and \mathbf{v}_i are the i^{th} price data and corresponding input data vector, respectively; and f_r is the regression operator. In this paper, the GPR approach is utilized for forecasting the electricity price and the motivations are four folds. ① GPR approach can give a possibility distribution for forecasted electricity prices, which cannot be accessed by other machine learning approaches. ② Prior knowledge, i.e., the indicators defined in Section II, can be added at this step and the shape of the regression model can be described by selecting different kernel functions [38]. ③ GPR approach is a statistical interpolation and non-parametric modeling approach based on Bayesian learning, which can provide a flexible hierarchical Bayesian framework for regression and prediction [39], [40]. ④ GPR approach is also an effective kernel-based machine learning approach and different kernel functions can be employed to deal with the nonlinearity of electricity prices. The precondition, i.e., assumption, of the usage of GPR approach is that the data should be in approximate Gaussian distribution, which is consistent with the actual situation of electricity price data for most situations.

To measure the relation between the forecasted electricity price z_{new} and the Gaussian distribution, the exponential kernel function, i.e., covariance function, is involved:

$$\varphi(\mathbf{v}_i, \mathbf{v}_{\text{new}}) = \sigma_f e^{-\frac{\|\mathbf{v}_i - \mathbf{v}_{\text{new}}\|}{2\sigma_l^2}} \quad (8)$$

where σ_f is the maximum allowable covariance; and σ_l is the length parameter. Thus, the matrices of kernel functions for training data, training data combined with testing data, and testing data can be respectively expressed as:

$$\Phi = \begin{bmatrix} \varphi(\mathbf{v}_1, \mathbf{v}_1) & \varphi(\mathbf{v}_1, \mathbf{v}_{\text{new}}) & \dots & \varphi(\mathbf{v}_1, \mathbf{v}_{N_{\text{train}}}) \\ \varphi(\mathbf{v}_2, \mathbf{v}_1) & \varphi(\mathbf{v}_2, \mathbf{v}_{\text{new}}) & \dots & \varphi(\mathbf{v}_2, \mathbf{v}_{N_{\text{train}}}) \\ \vdots & \vdots & & \vdots \\ \varphi(\mathbf{v}_{N_{\text{train}}}, \mathbf{v}_1) & \varphi(\mathbf{v}_{N_{\text{train}}}, \mathbf{v}_{\text{new}}) & \dots & \varphi(\mathbf{v}_{N_{\text{train}}}, \mathbf{v}_{N_{\text{train}}}) \end{bmatrix} \quad (9)$$

$$\Phi_* = [\varphi(\mathbf{v}_1, \mathbf{v}_{\text{new}}) \quad \varphi(\mathbf{v}_2, \mathbf{v}_{\text{new}}) \quad \dots \quad \varphi(\mathbf{v}_{N_{\text{train}}}, \mathbf{v}_{\text{new}})] \quad (10)$$

$$\Phi_{**} = \varphi(\mathbf{v}_{\text{new}}, \mathbf{v}_{\text{new}}) \quad (11)$$

It should be noted that several kernel functions can be selected for GPR approach, and the exponential kernel function can achieve the best performance, which will be further discussed in Section IV. In Gaussian processes (GPs), data located all over the domain are generated. In this way, any of the finite subset of the range of electricity price follows a

multivariate Gaussian distribution. Thus, the observations in the electricity price data set can always be regarded as a sample from a multivariate Gaussian distribution, i.e.,

$$\begin{bmatrix} z \\ z_{\text{new}} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Phi & \Phi^T \\ \Phi_* & \Phi_{**} \end{bmatrix} \right) \quad (12)$$

where $\mathbf{z} = [z_1, z_2, \dots, z_{N_{\text{train}}}]^T$. Furthermore, the possibility of z_{new} given \mathbf{z} is subjected to Gaussian distribution and can be derived into [38]:

$$z_{\text{new}} | \mathbf{z} \sim \mathcal{N}(\Phi_* \Phi^{-1} \mathbf{z}, \Phi_{**} - \Phi_* \Phi^{-1} \Phi_*^T) \quad (13)$$

It can be observed that the mathematical sense of $z_{\text{new}} | \mathbf{z}$ is exactly the electricity price forecasting result considering the training set, and the best electricity price forecasting result is the mean value of the distribution, i.e.,

$$f_{\text{GPR}}(\mathbf{v}_{\text{new}} | (\mathbf{v}_1, z_1), (\mathbf{v}_2, z_2), \dots, (\mathbf{v}_{N_{\text{train}}}, z_{N_{\text{train}}})) = \Phi_* \Phi^{-1} \mathbf{z} \quad (14)$$

where f_{GPR} is the GPR operator. It should be noted that the performance of the proposed GPR-based approach is greatly dependent on the setting of kernel function and its hyper-parameters, i.e., $\delta = (\sigma_f, \sigma_l)$. In fact, the best value of δ can be obtained when the prior probability $p(\delta | \mathbf{v}, \mathbf{z})$ reaches its maximum, and $\mathbf{v} = (\mathbf{v}_i)$. However, there is usually little prior knowledge about δ , so Bayes estimation [41], which is based on the maximum likelihood, is utilized to adjust δ by maximizing the post probability as:

$$\max \ln p(\mathbf{z} | \mathbf{v}, \delta) = -\frac{1}{2} \mathbf{z} \Phi^{-1} \mathbf{z} - \frac{1}{2} \ln \|\delta\| - \frac{N_{\text{train}}}{2} \ln 2\pi \quad (15)$$

Once the optimal hyper-parameters are determined, the electricity price can be forecasted accurately by (14). The flowchart of the proposed two-step day-ahead electricity price forecasting algorithm is given in Fig. 2. It should be noted that the proposed WKNN-based method and GPR-based approach are trained off-line.

IV. CASE STUDIES

To verify the effectiveness of the proposed WKNN-based method and GPR-based approach, the real data in the dominion area of the PJM electricity market from 2010/01/01 to 2019/12/31 are utilized, which can be accessed from PJM data miner [42]. The time interval between two data samples is 1 hour. Thus, nearly 88000 data samples are utilized and 80% of the available data are assigned as training data while the rest are assigned as testing data in this paper. Therefore, $N_{\text{train}} = 88000 \times 80\% = 70400$. All the case studies and comparisons are performed by MATLAB 2019a software deployed on Windows 10 platform with Intel Core i5-8250U and 8 GB RAM.

A. Verifications for Proposed WKNN-based Method

To evaluate the effectiveness of the proposed WKNN-based method, some evaluation indicators, i.e., accuracy, precision rate, recall rate, and F1 value of price spike detection are given as:

$$V_{\text{accuracy}}^{\text{spike}} = (N_{\text{TP}} + N_{\text{TN}}) / (N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}) \quad (16)$$

$$V_{\text{precision}}^{\text{spike}} = N_{\text{TP}} / (N_{\text{TP}} + N_{\text{FP}}) \quad (17)$$

$$V_{\text{recall}}^{\text{spike}} = N_{\text{TP}} / (N_{\text{TP}} + N_{\text{FN}}) \quad (18)$$

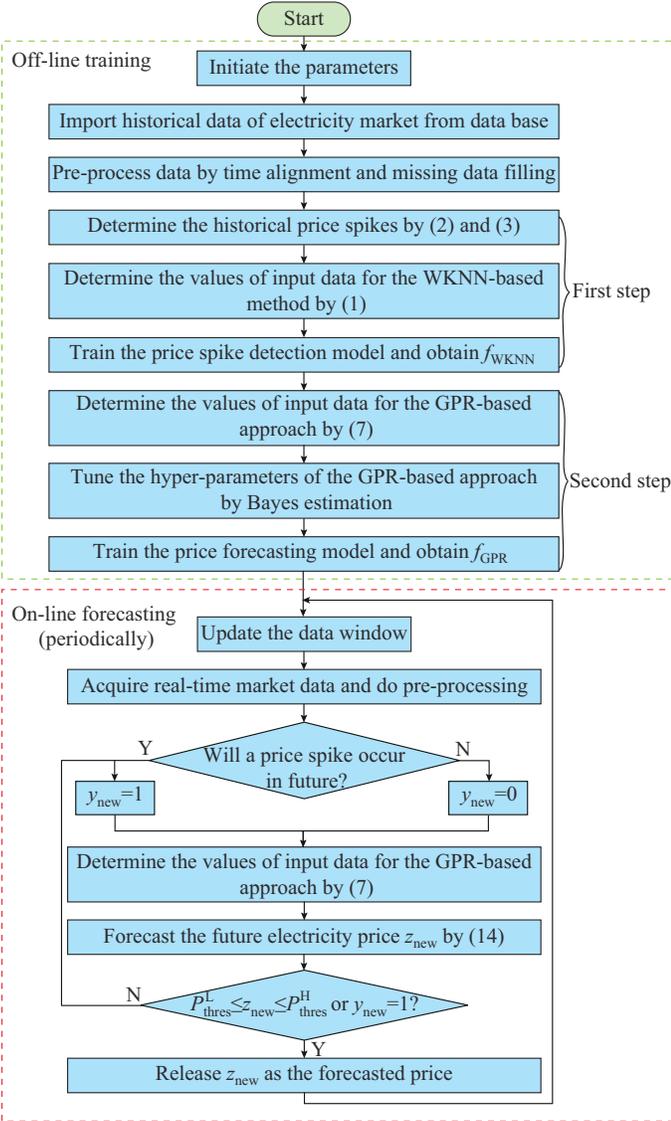


Fig. 2. Flowchart of proposed two-step day-ahead electricity price forecasting algorithm.

$$V_{F1}^{\text{spike}} = 2V_{\text{precision}}^{\text{spike}} V_{\text{recall}}^{\text{spike}} / (V_{\text{precision}}^{\text{spike}} + V_{\text{recall}}^{\text{spike}}) \quad (19)$$

where N_{FN} , N_{FP} , N_{TN} , and N_{TP} are the numbers of false negative, false positive, true negative, and true positive samples of price spikes, respectively. The larger the values of $V_{\text{accuracy}}^{\text{spike}}$, $V_{\text{precision}}^{\text{spike}}$, $V_{\text{recall}}^{\text{spike}}$, and V_{F1}^{spike} are, the better the detection method will be [43]. It is worth mentioning that the price spike detection is quite an unbalanced classification problem since the number of price spikes is much fewer than that of normal ones. Hence, the recall rate and F1 value are more essential for this work.

The overview of the electricity prices in the testing data set, i.e., 2018/01/01 to 2019/12/31 in dominion area of PJM, and the corresponding upper threshold $P_{\text{thres}}^H = 102.56$ \$/MWh and lower threshold $P_{\text{thres}}^L = -36.86$ \$/MWh determined based on historical data are shown in Fig. 3.

It can be observed from Fig. 3 that there are numerous price spikes and the accuracy of price forecasting will be largely influenced if they are not detected in advance. Thus, the proposed WKNN-based method is employed to detect

the price spikes and the results are shown in Table III. In fact, there are 412 price spikes and 16868 normal prices in total. It can be observed that the precision rate $V_{\text{precision}}^{\text{spike}}$ is 62.1% and the recall rate $V_{\text{recall}}^{\text{spike}}$ is 52.0%, which means that 62.1% of the detected price spikes are true spikes in practice and 52.0% of the true spikes are detected by the proposed WKNN-based method. Besides, its total accuracy $V_{\text{accuracy}}^{\text{spike}}$ is 97.7% and the F1 value V_{F1}^{spike} is 56.6%. The results obtained by the KNN method are also given in Table III, and it can be observed that the proposed WKNN-based method outperforms the KNN method in terms of all four indicators. Meanwhile, it can be observed from Table III that the proposed WKNN-based method achieves the largest values of $V_{\text{accuracy}}^{\text{spike}}$ and V_{F1}^{spike} among different methods. Although the recall rates of the methods based on the decision tree and bagged trees are higher than the proposed WKNN-based method, their precision rates are much smaller. Therefore, it can be concluded that the proposed WKNN-based method performs well in price spike detection and outperforms the other methods with regard to most of the evaluation indicators.

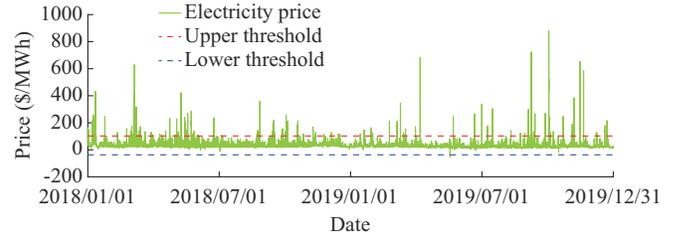


Fig. 3. Overview of electricity price from 2018/01/01 to 2019/12/31 in dominion area of PJM.

TABLE III
COMPARISONS AMONG DIFFERENT PRICE SPIKE DETECTION METHODS

Method	$V_{\text{precision}}^{\text{spike}}$ (%)	$V_{\text{recall}}^{\text{spike}}$ (%)	$V_{\text{accuracy}}^{\text{spike}}$ (%)	V_{F1}^{spike} (%)
Decision tree	54.1	52.7	97.7	53.4
Linear discriminant	55.3	18.8	93.3	28.1
Quadratic discriminant	49.0	23.6	95.0	31.9
Kernel Naïve Bayes	42.7	35.1	96.8	38.6
SVM	46.6	18.1	93.7	26.1
Bagged trees	27.2	52.8	97.7	35.9
RUSBoosted trees	66.0	47.6	97.5	55.3
KNN	51.2	49.4	97.6	50.3
Proposed	62.1	52.0	97.7	56.6

B. Verifications for Proposed GPR-based Approach

To evaluate the effectiveness of the proposed GPR-based approach, some evaluation indicators, i.e., root mean square error (RMSE) and mean absolute error (MAE) [44] are given as:

$$V_{\text{RMSE}} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} (\hat{z}_j - z_j)^2} \quad (20)$$

$$V_{\text{MAE}} = \frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} |z_j - \hat{z}_j| \quad (21)$$

where z_j and \hat{z}_j are the actual and forecasted values of elec-

tricity prices, respectively; and N_{test} is the total number of samples in the testing data set.

To better illustrate the effectiveness of the proposed GPR-based approach, the actual and forecasted electricity prices obtained by the proposed GPR-based approach in three typical days with and without price spikes are given in Figs. 4, 5, and 6.

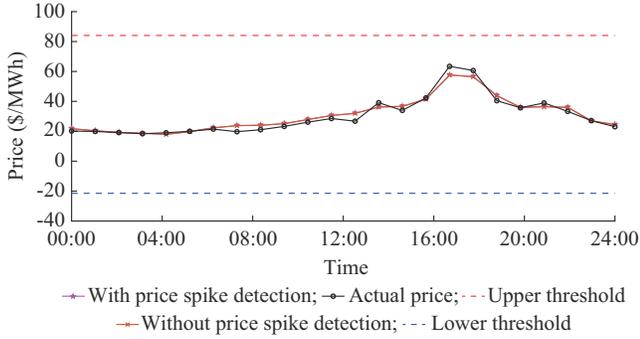


Fig. 4. Results obtained by proposed GPR-based approach on 2018/08/13 without price spikes.

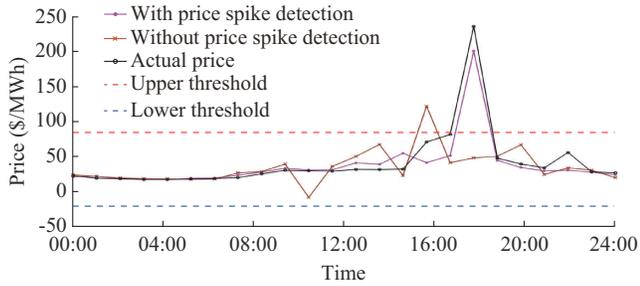


Fig. 5. Results obtained by proposed GPR-based approach on 2018/03/22 with a single price spike.

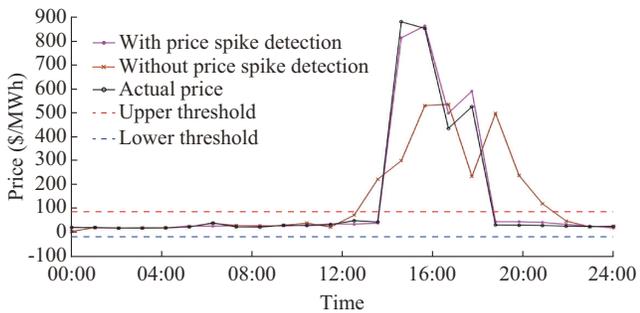


Fig. 6. Results obtained by proposed GPR-based approach on 2019/06/24 with multiple price spikes.

It can be observed from the figures that: ① the proposed GPR-based approach with the price spike detection in advance achieves quite smaller errors compared with the one without price spike detection; ② the performance difference between them is much larger in the days with a single price spike or multiple price spikes, which means that employing the price spike detection at first is very essential for price forecasting. For the whole testing data set, the proposed GPR-based approach also obtains a good performance with $V_{\text{RMSE}} = 15.862$ and $V_{\text{MAE}} = 4.806$. The results of comparisons among different approaches for electricity price forecasting

are also given in Table IV. It should be noted that the values of RMSE and MAE are related to the maximum absolute value (MAV) of prices in the market, and it would be quite smaller if per-unit values rather than actual values are used. Therefore, RMSE and MAE are only eligible for comparisons in the same market and at the same time.

TABLE IV
COMPARISONS AMONG DIFFERENT APPROACHES FOR ELECTRICITY PRICE FORECASTING WITH AND WITHOUT PRICE SPIKE DETECTION

Approach	Without price spike detection			With price spike detection		
	RMSE	MAE	Time (s)	RMSE	MAE	Time (s)
Interaction linear regression	21.497	8.283	7.023	17.437	6.162	5.624
Boosting trees	20.445	6.797	111.890	16.471	5.341	29.711
Exponential GPR	19.181	6.477	629.330	15.862	4.806	556.600

It can be observed from Table IV that the proposed GPR-based approach with exponential kernel, i. e., exponential GPR, achieves the smallest values of RMSE (15.862) and MAE (4.806) among all the approaches, which indicates that it outperforms other approaches concerning the forecasting accuracy. Among other approaches, the ensemble approach, i. e., boosting trees, achieves relatively good accuracy since it synthesizes more predictor structure or considers the time sequence characteristics. Besides, it can be observed that the results obtained by the approaches with price spike detection are much better than the ones without price spike detection, which means the first step of the proposed algorithm, i. e., price spike detection, is necessary and useful.

To demonstrate the performance of the proposed algorithm, several other state-of-the-art algorithms [6], [8], [45]-[48] are employed for comparisons, and the results are shown in Table V. It can be observed that the proposed algorithm can always achieve the least RMSE and MAE compared with other algorithms, which means the proposed algorithm obtains the best performance concerning forecasting accuracy.

TABLE V
COMPARISONS AMONG DIFFERENT ELECTRICITY PRICE FORECASTING ALGORITHMS

Algorithm	RMSE	MAE	Time (s)
Regression forest [6], [8]	17.851	5.210	11.918
RVM [45]	16.342	5.231	324.532
SVM [46]	17.142	5.435	194.232
CNN [47]	16.254	5.034	553.354
LSTM [48]	15.934	4.994	994.452
Proposed	15.862	4.806	556.600

As for the computation time, the proposed algorithm outperforms the CNN and LSTM algorithms, but is inferior to the other algorithms. Besides, the computation time of CNN and LSTM algorithms is relatively longer than the others.

The reason is that these algorithms utilize either the kernel and ensemble techniques or deep learning models, which would achieve better performance but cost a longer time than single machine learning algorithms. However, it should be mentioned that the computation time listed in Tables IV and V is associated with the off-line training stages of these algorithms, and the computation time used for online forecasting stages is very short, i.e., within 1 s; therefore, the proposed algorithm can meet the time requirements in practical use. Furthermore, the electricity price forecasting is day-ahead and will be just performed once a day commonly; therefore, it is worth acquiring a higher forecasting accuracy by a slightly longer computation time.

V. DISCUSSION

There are two critical parameters for the proposed algorithm, i.e., the parameter K in the proposed WKNN-based method and the kernel function utilized in the proposed GPR-based approach. Therefore, the analysis of the impact of these two parameter settings is performed in detail in this section.

A. Discussions About Parameter K in Proposed WKNN-based Method

As mentioned in Section II, the parameter K can be given in advance by experience or determined through the WKNN training process automatically [36]. To study the impact of setting of K on the proposed WKNN-based method, the corresponding sensitive analysis is performed and the results are shown in Table VI.

It can be observed that the setting of K does influence the effectiveness of price spike detection and the performance gets better gradually when $K < 4$ and gets worse gradually when $K > 4$. The best performance is obtained when $K = 4$ with the largest values of $V_{\text{precision}}^{\text{spike}}$, $V_{\text{recall}}^{\text{spike}}$, $V_{\text{accuracy}}^{\text{spike}}$, and $V_{\text{F1}}^{\text{spike}}$. Therefore, $K = 4$ is selected for price spike detection in this paper.

TABLE VI
SENSITIVE ANALYSIS OF PARAMETER K FOR PROPOSED WKNN-BASED METHOD

K	$V_{\text{precision}}^{\text{spike}}$ (%)	$V_{\text{recall}}^{\text{spike}}$ (%)	$V_{\text{accuracy}}^{\text{spike}}$ (%)	$V_{\text{F1}}^{\text{spike}}$ (%)
1	14.6	14.9	96.0	14.7
2	17.0	16.5	96.0	16.7
3	28.4	24.6	96.2	26.4
4	62.1	52.0	97.7	56.6
5	42.7	33.8	96.6	37.7
6	17.0	14.2	95.6	15.5
7	38.8	25.5	95.8	30.8
8	41.3	42.0	97.2	41.6
9	43.2	43.5	97.3	43.4
10	21.4	3.5	84.1	6.0

B. Discussions About Kernel Function of Proposed GPR-based Approach

As mentioned in Section III, the selection of kernel function for the proposed GPR-based approach is quite essential and has a large impact on the final performance of price forecasting. To be honest, there is no universal method for selecting the kernel function, and it is hard to give the specific reasons why selecting a given kernel function. In most situations, the kernel function is selected according to the engineers' experience and the trial-and-error method is also used usually to determine the kernel function with the best performance. If the performances of several kernel functions are similar, the kernel function with the least complexity should be used since it can avoid over-fitting problems and has a better generalization ability. Therefore, four commonly used kernel functions, i.e., squared exponential, Matern 5/2, rational quadratic, and exponential [49], [50], for the proposed GPR-based approach are compared and the results obtained by them are given in Table VII and shown in Figs. 7, 8, and 9. $r_d = \|\mathbf{v}_i - \mathbf{v}_{\text{new}}\|$ and α is a positive-valued scale-mixture parameter.

TABLE VII
COMPARISONS AMONG DIFFERENT KERNEL FUNCTIONS FOR PROPOSED GPR-BASED APPROACH

Kernel function	Formula $\varphi(\mathbf{v}_i, \mathbf{v}_{\text{new}})$	Without price spike detection			With price spike detection		
		RMSE	MAE	Time (s)	RMSE	MAE	Time (s)
Squared exponential	$\sigma_f^2 \exp\left(-\frac{r_d^2}{2\sigma_l^2}\right)$	20.723	7.838	829.160	23.387	6.471	2217.100
Matern 5/2	$\sigma_f^2 \left(1 + \frac{\sqrt{5} r_d}{\sigma_l} + \frac{5r_d^2}{3\sigma_l^2}\right)^{-\alpha} \exp\left(-\frac{\sqrt{5} r_d}{\sigma_l}\right)$	20.237	7.382	984.610	19.103	5.689	1021.900
Rational quadratic	$\sigma_f^2 \left(1 + \frac{r_d^2}{2\alpha\sigma_l^2}\right)^{-\alpha}$	20.445	7.683	1643.300	26.817	6.675	3150.800
Exponential	$\sigma_f \exp\left(-\frac{r_d}{2\sigma_l}\right)$	19.181	6.477	629.330	15.862	4.806	556.600

It can be observed that the exponential kernel function achieves the best performance with regard to all the values of RMSE, MAE, and computation time. Concretely, the difference is not very large in the typical day without a price

spike, while it is significant at the time of price spike in the typical day with a single price spike or multiple price spikes. Therefore, it can be concluded that the exponential kernel function is the most suitable one for the application

of electricity price forecasting of this paper.

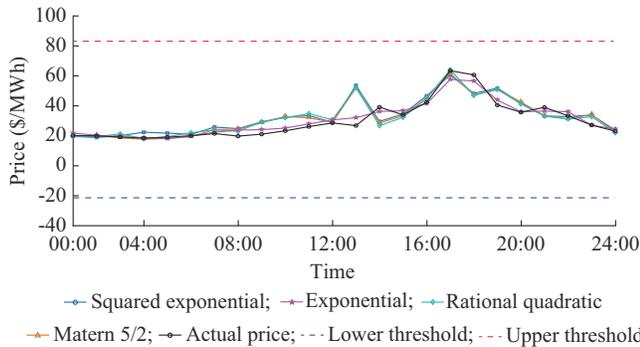


Fig. 7. Results obtained by proposed GPR-based approach with different kernel functions on 2018/08/13 without price spikes.

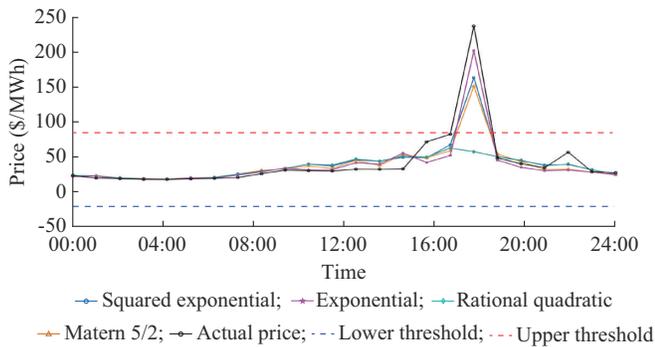


Fig. 8. Results obtained by proposed GPR-based approach with different kernel functions on 2018/03/22 with a single price spike.

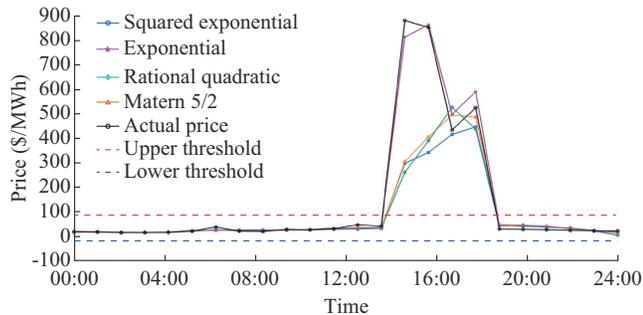


Fig. 9. Results obtained by proposed GPR-based approach with different kernel functions on 2019/06/24 with multiple price spikes.

VI. CONCLUSION

In this paper, a two-step day-ahead electricity price forecasting algorithm is proposed, in which the price spikes are detected by the WKNN-based method first and then the electricity price is forecasted by the GPR-based approach. The necessity and effectiveness of price spike detection before forecasting are systematically addressed, and comparisons with several other price forecasting algorithms are carried out. Besides, the selection of the kernel function for the proposed GPR-based approach is also discussed. The following conclusions are attained.

1) It is essential to detect the price spikes at first and utilize the spike indicator as one of the features for the proposed GPR-based approach. Simulation results show that the

price forecasting algorithm with the price spike detection makes much smaller errors compared with the one without the price spike detection.

2) The performances of the proposed GPR-based approach are closely related to its kernel function and the exponential kernel function is demonstrated as the one that can obtain better performance compared with other kernel functions after detailed comparisons.

3) The performances of price forecasting algorithms are greatly influenced by the occurrence of price spikes. The number of forecasting errors is much larger in the days with multiple price spikes than the days without price spikes. Therefore, the more accurate price spike detection is the key research point of electricity price forecasting, which is also part of our future work.

REFERENCES

- [1] Z. Zhang, R. Li, and F. Li, "A novel peer-to-peer local electricity market for joint trading of energy and uncertainty," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1205-1215, Mar. 2020.
- [2] Y. Du, F. Li, H. Zandi *et al.*, "Approximating Nash equilibrium in day-ahead electricity market bidding with multi-agent deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 3, pp. 534-544, May 2021.
- [3] P. Razmi, M. O. Buygi, and M. Esmalifalak, "A machine learning approach for collusion detection in electricity markets based on Nash equilibrium theory," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 1, pp. 170-180, Jan. 2021.
- [4] S. Shenoy and D. Gorinevsky, "Data-driven stochastic pricing and application to electricity market," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 6, pp. 1029-1039, Sept. 2016.
- [5] A. Radovanovic, T. Nesti, and B. Chen, "A holistic approach to forecasting wholesale energy market prices," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4317-4328, Nov. 2019.
- [6] J. Lago, F. D. Ridder, and B. D. Schutter, "Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy*, vol. 221, pp. 386-405, Feb. 2018.
- [7] P. Rana, J. Vilar, and G. Aneiros, "On the use of functional additive models for electricity demand and price prediction," *IEEE Access*, vol. 6, pp. 9603-9613, Mar. 2018.
- [8] I. Juárez, J. Mira-McWilliams, and C. González, "Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, bagging and random forests," *IET Generation, Transmission & Distribution*, vol. 9, no. 11, pp. 1120-1128, Jun. 2015.
- [9] K. Wang, C. Xu, Y. Zhang *et al.*, "Robust big data analytics for electricity price forecasting in the smart grid," *IEEE Transactions on Big Data*, vol. 5, no. 1, pp. 34-45, Jan. 2019.
- [10] P. Mandal, T. Senjyu, N. Urasaki *et al.*, "A novel approach to forecast electricity price for PJM using neural network and similar days method," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2058-2065, Nov. 2007.
- [11] N. M. Pindoriya, S. N. Singh, and S. K. Singh, "An adaptive wavelet neural network-based energy price forecasting in electricity markets," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1423-1432, Aug. 2008.
- [12] X. Chen, Z. Y. Dong, K. Meng *et al.*, "Electricity price forecasting with extreme learning machine and bootstrapping," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2055-2062, Nov. 2012.
- [13] C. Wan, M. Niu, Y. Song *et al.*, "Pareto optimal prediction intervals of electricity price," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 817-819, Jan. 2017.
- [14] C. Wan, Z. Xu, Y. Wang *et al.*, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 463-470, Jan. 2014.
- [15] M. Alamaniotis, D. Bargiotas, N. G. Bourbakis *et al.*, "Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2997-3005, Nov. 2015.
- [16] O. Abedinia, N. Amjadi, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems,"

- IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62-74, Jan. 2017.
- [17] A. T. Lora, J. M. R. Santos, A. G. Exposito *et al.*, "Electricity market price forecasting based on weighted nearest neighbors techniques," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1294-1301, Aug. 2007.
- [18] S. Ambazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using recurrent neural network," *IEEE Systems Journal*, vol. 7, no. 4, pp. 866-872, Dec. 2013.
- [19] L. Wang, Z. Zhang and J. Chen, "Short-term electricity price forecasting with stacked denoising autoencoders," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2673-2681, Jul. 2017.
- [20] D. H. Vu, K. M. Muttaqi, A. P. Agalgaonkar *et al.*, "Short-term forecasting of electricity spot prices containing random spikes using a time-varying autoregressive model combined with kernel regression," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5378-5388, Sept. 2019.
- [21] H. Chitsaz, P. Zamani-Dehkordi, H. Zareipour *et al.*, "Electricity price forecasting for operational scheduling of behind-the-meter storage systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6612-6622, Nov. 2018.
- [22] R. Weron, "Electricity price forecasting: a review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030-1081, Oct. 2014.
- [23] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: a review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, vol. 81, no. 1, pp. 1548-1568, Jan. 2018.
- [24] J. Zhao, Z. Y. Dong, X. Li *et al.*, "A framework for electricity price spike analysis with advanced data mining methods," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 376-385, Feb. 2007.
- [25] J. Zhao, Z. Y. Dong, and X. Li, "Electricity market price spike forecasting and decision making," *IET Generation, Transmission & Distribution*, vol. 1, no. 4, pp. 647-654, Jul. 2007.
- [26] H. Manner, D. Türk, and M. Eichler, "Modeling and forecasting multivariate electricity price spikes," *Energy Economics*, vol. 60, pp. 255-265, Nov. 2016.
- [27] H. S. Sandhu, L. Fang, and L. Guan, "Forecasting day-ahead price spikes for the Ontario electricity market," *Electric Power Systems Research*, vol. 141, pp. 450-459, Dec. 2016.
- [28] L. M. Sixel. (2019, Aug.). New price adders boosting electricity prices in Texas. [Online]. Available: <https://www.chron.com/business/energy/article/New-price-adders-boosting-electricity-prices-in-14299363.php?c=mpid=ffcp>
- [29] Z. Jing, J. Zhu, and R. Hu, "Sizing optimization for island microgrid with pumped storage system considering demand response," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 4, pp. 791-801, Jul. 2018.
- [30] Pennsylvania-New Jersey-Maryland Interconnection (PJM). (2021, Jan.). Historical load forecasts. [Online]. Available: https://dataminer2.pjm.com/feed/load_frcstd_hist/definition
- [31] PJM. (2021, Jan.). Scheduled generation. [Online]. Available: https://dataminer2.pjm.com/feed/rt_and_self_ecomax/definition
- [32] PJM. (2021, Jan.). Daily generation capacity. [Online]. Available: https://dataminer2.pjm.com/feed/day_gen_capacity/definition
- [33] Pennsylvania-New Jersey-Maryland Interconnection (PJM). (2021, Jan.). Forecasted generation outages. [Online]. Available: https://dataminer2.pjm.com/feed/frcstd_gen_outages/definition
- [34] Pennsylvania-New Jersey-Maryland Interconnection (PJM). (2021, Jan.). Operations summary-actual operational statistics. [Online]. Available: https://dataminer2.pjm.com/feed/ops_sum_prev_period/definition
- [35] Pennsylvania-New Jersey-Maryland Interconnection (PJM). (2021, Jan.). Generation and extra high voltage losses. [Online]. Available: https://dataminer2.pjm.com/feed/gen_chv_losses/definition
- [36] J. Song, J. Zhao, F. Dong *et al.*, "A novel regression modeling method for PMSLM structural design optimization using a distance-weighted KNN algorithm," *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 4198-4206, Sept. 2018.
- [37] W. Zuo, D. Zhang, and K. Wang, "On kernel difference-weighted k -nearest neighbor classification," *Pattern Analysis and Applications*, vol. 11, pp. 247-257, Jan. 2008.
- [38] M. Ebdn. (2020, Jan.). Gaussian processes for regression: a quick introduction. [Online]. Available: <https://arxiv.org/pdf/1505.02965.pdf>
- [39] J. Han, X. Zhang, and F. Wang, "Gaussian process regression stochastic volatility model for financial time series," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 6, pp. 1015-1028, Sept. 2016.
- [40] J. Feng, X. Jia, H. Cai *et al.*, "Cross trajectory Gaussian process regression model for battery health prediction," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 5, pp. 1217-1226, Sept. 2021.
- [41] B. Dulek, "A restricted Bayes approach to joint detection and estimation under prior uncertainty," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1767-1782, Feb. 2018.
- [42] PJM. (2021, May). PJM data miner 2. [Online]. Available: <https://dataminer2.pjm.com/list>
- [43] S. Liu, S. You, Z. Lin *et al.*, "Data-driven event identification in the U.S. power systems based on 2D-OLPP and RUSBoosted trees," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 94-105, Jan. 2022.
- [44] N. R. Draper and H. Smith, *Applied Regression Analysis*. Hoboken: Wiley-Interscience Publication, 1998.
- [45] M. Alamaniotis, D. Bargiotas, N. G. Bourbakis *et al.*, "Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2997-3005, Nov. 2015.
- [46] L. M. Saini, S. K. Aggarwal, and A. Kumar, "Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in national electricity market," *IET Generation, Transmission & Distribution*, vol. 4, no. 1, pp. 36-49, Jan. 2010.
- [47] Y. Hong, J. V. Taylor, and A. C. Fajardo, "Locational marginal price forecasting using deep learning network optimized by mapping-based genetic algorithm," *IEEE Access*, vol. 8, pp. 91975-91988, May 2020.
- [48] S. Zhou, L. Zhou, M. Mao *et al.*, "An optimized heterogeneous structure LSTM network for electricity price forecasting," *IEEE Access*, vol. 7, pp. 108161-108173, Aug. 2019.
- [49] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge: MIT Press, 2006.
- [50] R. M. Neal, *Bayesian Learning for Neural Networks*. New York: Springer, 1996.

Shengyuan Liu received the B.E. degree in electrical engineering from Shandong University, Jinan, China, in 2017. He is currently pursuing the Ph.D. degree at the College of Electrical Engineering, Zhejiang University, Hangzhou, China. He is also a visiting Ph.D. student at The University of Tennessee, Knoxville, USA, for one year from 2019 to 2020. His research interests include wide-area monitoring and control of power systems, situation awareness of power systems, and data-driven approaches in power systems.

Yicheng Jiang received the B.E. degree from South China University of Technology, Guangzhou, China, in 2017, and the M.E. degree from Zhejiang University, Hangzhou, China, in 2020, both in electrical engineering. She is currently working at State Grid Zhejiang Electric Power Corporation, Hangzhou, China. Her research interests include power system control, electricity market, and renewable energy.

Zhenzhi Lin received the Ph.D. degree in electrical engineering from South China University of Technology, Guangzhou, China, in 2008. He was a Research Assistant at the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, China, from 2007 to 2008, a Research Scholar at the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, USA, from 2010 to 2011, and a Research Associate at the College of Engineering and Computing Sciences, Durham University, Durham, UK, from 2013 to 2014. He is currently a Professor at the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include wide-area monitoring and control of power systems, controlled islanding and power system restoration, and data mining in power systems.

Fushuan Wen received the B.E. and M.E. degrees from Tianjin University, Tianjin, China, in 1985 and 1988, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1991. He joined the faculty of Zhejiang University in 1991, and has been a Full Professor in electrical engineering since 1997. He had been a University Distinguished Professor in electrical engineering, the Deputy Dean of the School of Electrical Engineering and the Director of the Institute of Power Economics and Electricity Markets at South China University of Technology, Guangzhou, China, from 2005 to 2009. He is a Professor in electrical engineering with the Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology, Tallinn, Estonia, taking leave from Zhejiang University. He is the Editor-in-Chief of IET Energy Conversion and Economics, and a Deputy Editor-in-Chief of Automation of Electric Power Systems. He also serves as the Editor, Subject Editor, and Associate Editor of a few international journals. His research interests include power industry restructuring,

power system alarm processing, fault diagnosis and restoration strategies, as well as smart grids and electric vehicles.

Yi Ding received the B.Eng. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2007, both in electrical engineering. He is a Professor at the College of Electrical Engineering, Zhejiang University, Hangzhou, China. Before joining Zhejiang University, he was an Associate Professor at the Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark. He also held research and teaching positions at University of Alberta, Edmonton, Canada and Nanyang Technological University. He was a Consultant as Energy Economist for Asian Development Bank in 2010. He is editorial member of international journals of

Electric Power System Research and Journal of Modern Power Systems and Clean Energy. He is also a Guest Editor for the special section of IEEE Transactions on Power Systems. He is a member of IEC working groups for micro-grid standards. His research areas include power system planning and reliability evaluation, smart grid, and complex system risk assessment.

Li Yang received the Ph.D. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2004. She held a Postdoctoral position at the Department of Electrical Engineering, Turin Polytechnic University, Turin, Italy, from 2007 to 2008. She is currently an Associate Professor with the College of Electrical Engineering, Zhejiang University. Her research interests include electricity market, power system economics, and distribution network planning.